A Transport Coding Gain Estimation in the Conditions of Time Limitation for Maximum Acceptable Message Delay

Evgenii Krouk[1], Anton Sergeev[2] and Mikhail Afanasev[3]

[1]National Research University Higher School of Economics,
Moscow Institute of Electronics and Mathematics (MIEM HSE)
Tallinskaya Str. 34, 123458, Moscow
1)ekrouk@hse.ru
[2,3]Saint-Petersburg State University of Aerospace Instrumentation,
B.Morskaya 67, 190000, Saint-Petersburg, Russia
2)slaros@k36.org,3)af-mm@yandex.ru

Abstract. Recently, low message delay becomes more and more important. It allows an implementing of new applications and services interacting in almost real time that was impossible before. A requirement of low message delay now is a part of many communication standards such as 5G and recommendation papers of future standards such as Tactile Internet proposed by ITU-T. In this research we perform an efficiency estimation of the transport coding for packets with an exponentially distributed delay. We consider a traffic model, that is critical to message delay, and an influence of the transport coding on such traffic types and on the message delay jitter. Our research shows that the transport coding can be used not only for a decreasing of the average message delay but and for a decreasing of the message delay jitter.

Keywords: transport coding, latency, delay, lattency critical applications, error correcting coding, network coding, Kleinrock's model

1 Introduction

There are types of applications and services that are critical to message delay. For instance, VoIP, remote control systems, spatially distributed security systems, etc. These systems use audio, video and control commands. When something happens these systems should response immediately in order to prevent possible consequences or support suitable quality of service. We can underline several parameters of such systems:

1. t_{max} is the maximum acceptable message delay (e.g. 250 ms for VoIP [1]);
2. Message delay jitter is a difference between average message delay and maximum and minimum message delay: $t'_{avg} - t'_{min}, t'_{max} - t'_{avg}$ [1].

Let us consider application, tranport and network levels of OSI. Usual approaches to decrease message delay are:

1. Adjusting packet routing procedures in the network;
2. Adjusting a packet prioritization in the network (QoS);
3. Using buffers in order to fight with the message delay jitter;
4. etc.

An alternative approach is the transport coding first proposed by Krouk, E.A., and Kabatiansky, G.A., in [2][3]. The main purpose of the transport coding is a decreasing of the average message delay in the network. But it also allows to organize a prioritization of urgent messages [4].

In initial research the simplified Kleinrock's network model was considered. In subsequent research, some simplified assumptions were dropped in order to consider more realistic cases, like non-exponential packet delay [5], different channel capacities [6], non-heterogenious network structure [7], delivery of a message during limited time [8], application of the transport coding [9].

The difference of this research from others about the transport coding is that we add a new realistic assumption of the network model: maximum time limit of message delivery and the volume of messages that we can afford to loose. We perform an efficiency analysis of the transport coding for packets with exponentially distributed delay in the network with the added realistic assumption.

First we need to consider a network model.

2 The Transport Coding

2.1 The Network Model

In [2], Kleinrock's network model [10] is considered with additional assumptions.

The network consists of M channels and N nodes. All channels are absolutely reliable and do not make errors. They have the same channel capacities C. Nodes perform a processing of received packets, including routing procedures, storing packets, controlling of packet queues. Nodes are absolutely reliable and do not make errors. The time of processing at nodes is performed instantly and is not taken into account in the consideration.

Packets arriving in the ith network node and moving towards to the jth network node form a Poisson process with the equal average value γ_{00} (packets per seconds) for every pair of i and j. The value

$$\gamma_0 = N\gamma_{00}$$

is an external traffic of the ith node, and the value

$$\gamma = N\gamma_0 = N^2\gamma_{00}$$

is the total external traffic of the network.

The network has buffers for packets with infinite capacity. It means that packets in the network can never be lost bacause of buffer overflow.

All messages arriving in the network consist of k packets with length m. The length of packets inside the network satisfies an assumption of independence [10]: every time when a packet has been received at a node, a new length of this packet b' is chosen independently with pdf

$$P(b) = \mu e^{-\mu b}, b \geq 0,$$

where b (the size of a packet in bits) is a parameter of the pdf function.

The packet flow going through the channels is a Poisson process with the equal average value λ_0 (packets per a second) for every channel of the network. The total internal traffic in the network is

$$\lambda = M\lambda_0.$$

Let us denote ρ as the value of channel load in the network, then the average packet delay is:

$$\bar{t}(\rho) = \frac{\lambda}{\mu C \gamma} \frac{1}{1-\rho}. \tag{1}$$

where ρ is

$$\rho = \frac{\lambda_0}{\mu C}.$$

Let us make two additional assumptions that are not contained in the original Kleinrock's network model:

1. Packet delays in the network are random and independent values distributed according to an exponential distribution law with the average value $\bar{t}(\rho)$

   $$F_\rho(t) = 1 - e^{-t/\bar{t}(\rho)}, t \geq 0, \tag{2}$$

   where t (time units) is a parameter of the cdf function. An exponential distribution is used for describing of data flows in many researches, in particular in [11].
2. The routing procedure in the network is chosen in such a way that an increasing of an external traffic leads to the uniform increasing of an internal traffic of each channel.

2.2 The Message Encoding Procedure

In [2], the following procedure named as the transport coding was proposed: Initially, original packets arriving in the network have the length m. Let us consider each character of a packet as an element of $GF(2^m)$. It allows us to encode each message by a 2^m MDS code (e.g. a Reed-Solomon code). Then each original message consisting of k packets maps to an encoded message consisting of n packets. After encoding procedure, we will send the encoded message instead of the original message. It leads to an increasing of the network load ρ by a factor $n/k = 1/R$ and the average packet delivery time becomes $\overline{t}(\rho/R)$.

It is well known that the word of MDS(n, k) code can be decoded by any k from n information characters. Hence, k first arrived encoded packets are enough to obtain our original message.

The message delay of original messages equals to the kth order statistic from $k - \overline{T}_{k:k}$. While the message delay of encoded messages equals to the kth order statistic from $n - \overline{T}_{k:n}$.

In the case of when m is too long, we need to deal with big fields GF, that is not desirable because of processing difficulty and it is unacceptable for low power devices like Internet Of Things. In [2], the solution of this case is described.

3 The Problem Statement

Perform efficiency analysis of the transport coding for packets with an exponentially distributed packet delay (2) in the network model described above and traffic types that are critical to message delay.

Let us denote the direct problem statement and the inverse problem statement:

1. The direct problem statement: minimize the delivery time of packets arrived later than t_{max};
2. The inverse problem statement: find optimal code parameters when t_{max} is fixed value.

3.1 The Traffic Model That Is Critical to Message Delay

Network delay is one of the main factors which can degrade the Quality of Experience (QoE) of network services [12]. So to prevent the degradation of the perceived quality of the services with delay constraints, a maximum limit is defined in the most of communication systems and protocols. This applies both real-time (VoIP, RDP, gaming, remote control, video [13] and audio [14] streaming) and non real-time services (instant messaging, M2M metering etc.). A guidelines for the maximum allowed latency and proposed multiplexing periods for different services/scenarios can be found for example in [15] and [16]. Now there are many research papers considering delay limit issues for video surveillance, real-time traffic over TCP [17] and other applications. Servide latency limits are in the focus of 3GPP 5G recommendations. In particular 3GPP Technical specification "Mission Critical Video over LTE" [18] defines end-to-end maximum latency for different scenarios: urgent real time video transmissions, Robots video remote control etc.

We consider this fact in our research and network model and show that the transport coding can be used both for significant decreasing of the average message delay and jitter.

First we start with a traffic model which can be described by the following parameters:

1. t_{max} is the maximum acceptable packet delivery time;
2. $p = Pr\{\overline{T}_{k:n} > t_{max}\}$ is the volume of messages that have arrived later than t_{max};
3. k is the amount of packets of an uncoded message.

Taking into account the assumption (2) for considered network model, packet delivery time is a random variable distributed according to an exponential distribution law. For this traffic model, it is impossible to talk about methods that can guarantee that all messages will be delivered for the time less than t_{max}, but we can talk about a minimizing of the value p.

3.2 The Average Packet Delivery Time

The average packet delivery time of an individual packet (1) depends on many parameters, such as the total intensity of the internal traffic λ, the total intensity of the external traffic γ, μ, the capacity of all channels C. We can see that when $\rho = 0$ then $\frac{\lambda}{\gamma \mu C}$ is an initial packet delivery time in the empty network, and $\frac{1}{1-\rho}$ is the increasing coefficient depending only on ρ. Thus, absolute values of these parameters are not important. Only the initial packet delivery time in the empty network is of interest. Let us denote it as $\bar{t}_s$ and rewrite (1) as

$$\bar{t}(\rho) = \bar{t}_s \frac{1}{1-\rho}. \tag{3}$$

For all next calculations we consider $\bar{t}_s = 2$. This value was taken thus that for $t_{max} = 9$ the considered network model will not be able to deliver messages for the average time less than t_{max} with the intensity of the channel traffic ρ. We need to note that using time units are abstract values and they can be expressed in needed units for a particular task.

3.3 The Direct Problem Statement

Let us fix the value t_{max}, what is the probability $p(\rho)$ of that the message will be delivered later than t_{max} when ρ is known?

For the next calculations let us set $k = 8$.

$$\bar{t}(\rho/R) = \bar{t}(\rho) \frac{1-\rho}{1-\frac{\rho}{R}}$$

$$p = 1 - Pr\{\overline{T}_{k:n} \le t_{max}\} = 1 - \sum_{i=k}^{n} C_n^i (1-e^a)^i e^{a(n-i)}, n = \frac{k}{R}$$

where

$$a = -\frac{t_{max}}{\bar{t}(\rho/R)} = -\frac{t_{max}(1-\frac{\rho}{R})}{t(\rho)(1-\rho)}$$

In Fig. 1, we can see the probability of message lost for different code rates for the case when $t_{max} = 9$ and the intensity of channel load $\rho = 0.6$:

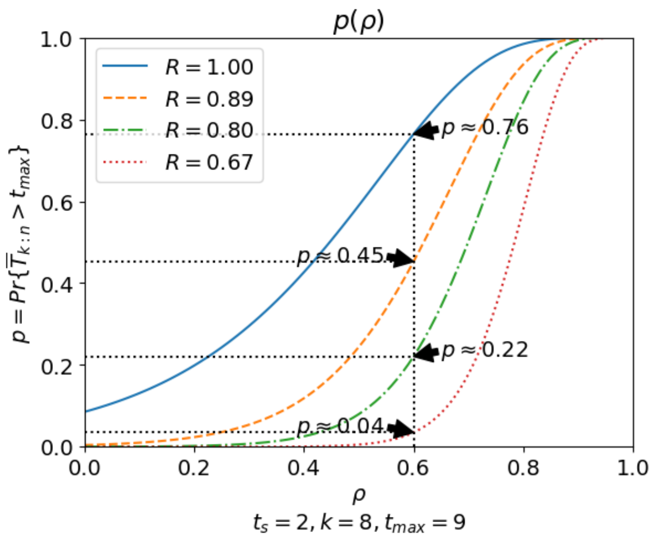

Fig. 1. The direct problem statement when $t_{max} = 9$

1. In the case without the transport coding ($R = 1$) the probability of message lost is $p \approx 0.76$;
2. In the case $R = 0.89, p = 0.45$;
3. In the case $R = 0.80, p = 0.22$;
4. In the case $R = 0.67, p = 0.04$.

When $p = 0.2$, without the transport coding the maximum channel load is $\rho_{max} \approx 0.2$, but in the case with transport coding and the code rate $R = 0.8$, the maximum channel load can be increased to $\rho_{max} \approx 0.6$.

3.4 The Influence of the Transport Coding on the Message Delay Jitter

First we need to start with determining a dispersion and a standard deviation for the average message delay. Taking into account the assumption (2) about an exponential distribution law of packet delay, we can obtain:

$$M[T] = \bar{t}(\rho)$$

$$D[T] = \bar{t}(\rho)^2$$

The dispersion of the kth order static [19]:

$$D[T_{k:n}] = D[T] \sum_{i=n-k+1}^{n} i^{-2} = \bar{t}(\rho \backslash R)^2 \sum_{i=n-k+1}^{n} i^{-2}$$

$$D[T_{k:k}] = D[T]\sum_{i=1}^{k} i^{-2} = \bar{t}(\rho)^2 \sum_{i=1}^{k} i^{-2}$$

The gain in the standard deviation is

$$f(R) = \frac{\sigma[T_{k:k}]}{\sigma[T_{k:n}]} = \frac{\sqrt{D[T_{k:k}]}}{\sqrt{D[T_{k:n}]}} = \frac{\bar{t}(\rho)\sqrt{\sum_{i=1}^{k} i^{-2}}}{\bar{t}(\rho\backslash R)\sqrt{\sum_{i=n-k+1}^{n} i^{-2}}} = \frac{(1-\rho\backslash R)\sqrt{\sum_{i=1}^{k} i^{-2}}}{(1-\rho)\sqrt{\sum_{i=n-k+1}^{n} i^{-2}}}$$

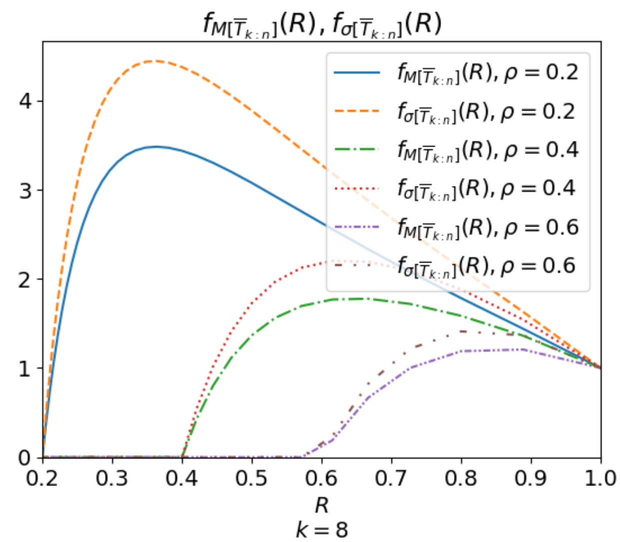

Fig. 2. The gain in the standard deviation

In Fig. 2, we can see a curve line of the gain in the standard deviation depending on the intensity of channel load and the assessment of the gain in the standard deviation obtained in [2]. The following cases were depicted in that figure:

1. $\sigma[\overline{T}_{k:n}]$ when $\rho = 0.2$;
2. $M[\overline{T}_{k:n}]$ when $\rho = 0.2$;
3. $\sigma[\overline{T}_{k:n}]$ when $\rho = 0.4$;
4. $M[\overline{T}_{k:n}]$ when $\rho = 0.4$;
5. $\sigma[\overline{T}_{k:n}]$ when $\rho = 0.6$;
6. $M[\overline{T}_{k:n}]$ when $\rho = 0.6$.

We can conclude from obtained plots that the transport coding can decrease the standard deviation and it means that the transport coding can also decrease jitter of message delay. In particular, in the case when $\rho = 0.2$ and $R = 0.36$, the gain of the standard deviation is ≈ 4.5 times, while the gain of the average message delay [2] is ≈ 3.5 times.

3.5 The Inverse Problem Statement

Let us fix the value t_{max} and the probability p, that the message has arrived later than t_{max}. What is a code rate we need to use in order to keep chosen $p(\rho)$?

$$p \geq 1 - \sum_{i=k}^{n} C_n^i (1 - e^a)^i e^{a(n-i)}, n = k \dots n_{max}$$

where n_{max} is a maximum amount of packets from which our message consists of, and

$$a = -\frac{t_{max}}{\bar{t}(\rho/R)} = -\frac{t_{max}(1 - \frac{\rho}{R})}{t(\rho)(1 - \rho)}$$

For next calculations let us set $n_{max} = 100$.

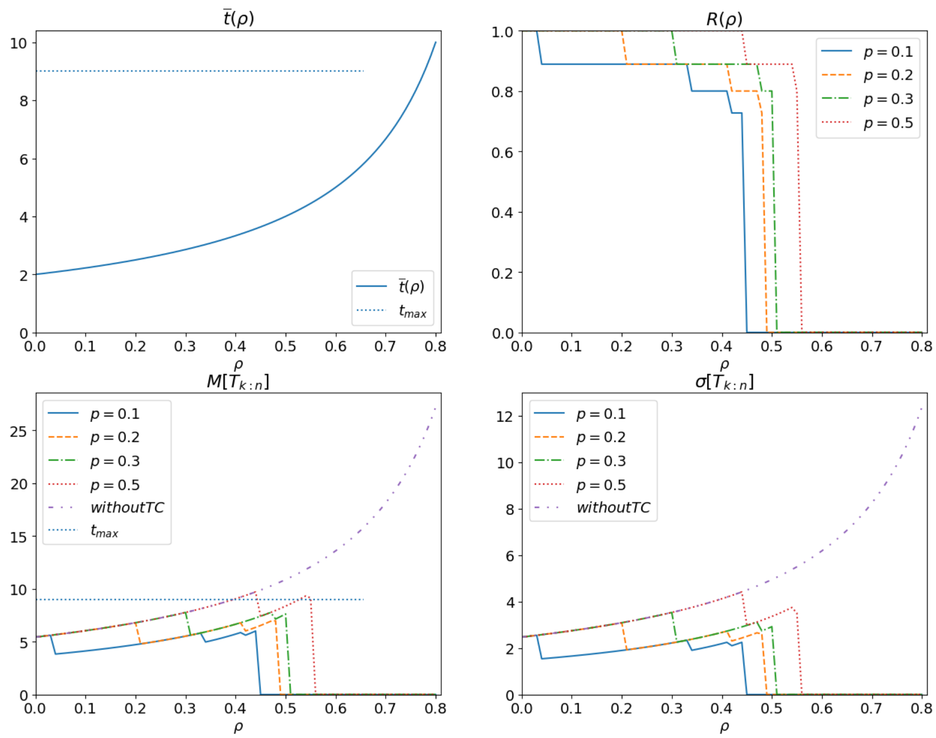

Fig. 3. The inverse problem statement when $t_{max} = 9$

In Fig. 3, the following curves were depicted for the case $t_{max} = 9$:

1. t_{max};
2. $\bar{t}(\rho)$;
3. $M[\overline{T}_{k:n}](\rho)$, $\sigma[\overline{T}_{k:n}](\rho)$, $R(\rho)$ when $p = 0.1$;
4. $M[\overline{T}_{k:n}](\rho)$, $\sigma[\overline{T}_{k:n}](\rho)$, $R(\rho)$ when $p = 0.2$;
5. $M[\overline{T}_{k:n}](\rho)$, $\sigma[\overline{T}_{k:n}](\rho)$, $R(\rho)$ when $p = 0.5$;
6. $M[\overline{T}_{k:n}](\rho)$ when $R = 1$;
7. $\sigma[\overline{T}_{k:n}](\rho)$ when $R = 1$.

We can conclude that when the channel load ρ increases, the code rate decreases too in order to maintain chosen parameters of the traffic model, and then after some ρ_{max}, the transport coding will not be able to maintain chosen parameters.

Conclusion

In this research we suggested new traffic model that is critical to the delay. The traffic model consists of the maximum acceptable message delay t_{max} and the probability that a message has been delivered later than $t_{max}, p = Pr\{\overline{T}_{k:n} > t_{max}\}$. We performed efficiency analysis of the transport coding for considered traffic model.

For considered network model parameters ($t_s = 2$) and the traffic model ($t_{max} = 9, k = 8, n_{max} = 100$), we concluded that the transport coding allows:

1. Decreasing of the message delay jitter. In the case of $\rho = 0.2$ and $R = 0.36$, the gain of the standard deviation is ≈ 4.5 times, while the gain of the average message delay [2] is ≈ 3.5 times;
2. Increasing of the maximum intensity of channel load ρ_{max} in order to maintain chosen parameters of traffic and network models. In the case of $p = 0.2$ and without transport coding, we can maintain chosen parameters of channel and network models with the maximum intensity of channel load is $\rho_{max} \approx 0.2$, while in the case with transport coding and the code rate $R = 0.8$, this parameter can be increased to $\rho_{max} \approx 0.6$;
3. Decreasing of the probability p of that the message has arrived later than t_{max} in the case of a fixed intensity of the channel load ρ. In the case without the transport coding when $\rho = 0.6$ the probability of message lost is $p = 0.76$, while the transport coding allows decreasing of this value to $p = 0.04$ with a coderate $R = 0.67$.

Further possible research directions of the transport coding can include:

1. Efficency analysis of the transport coding in more realistic cases including parameters such as: MTU size, packet's length in bytes, packet delay in milliseconds, etc.;
2. Considering of real protocols that are critical to the message delay.

References

1. Swale, R.: Voice over IP: systems and solutions. let (2001)
2. Kabatiansky, G., Krouk, E., Semenov, S.: Error correcting coding and security for data networks: analysis of the superchannel concept. John Wiley & Sons (2005)
3. Kabatianskii, G., Krouk, E.: Coding decreases delay of messages in networks. IEEE International Symposium on Information Theory. Proceedings (1993)
4. Krouk, E., Semenov, S.: Transmission of priority messages with the help of transport coding. Telecommunications, 2003. ICT 2003. 10th International Conference on 2, 1273–1277 (2003)
5. Malichenko, D., Krouk, E.: Estimation of the mean message delay for transport coding. Intelligent Interactive Multimedia Systems and Services pp. 239–249 (2015)
6. Malichenko, D.: Efficiency of transport layer coding in networks with changing capacities. Problems of Redundancy in Information and Control Systems (REDUNDANCY), 2016 XV International Symposium pp. 82–86 (2016)
7. Malichenko, D.: Transport layer coding with variable coding rate. Problems of Redundancy in Information and Control Systems (REDUNDANCY), 2014 XIV International Symposium on pp. 71–73 (2014)
8. Krouk, E., Semenov, S.: Delivery of a message during limited time with the help of transport coding. Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on pp. 1–5 (2004)
9. Krouk, E., Semenov, S.: Application of coding at the network transport level to decrease the message delay. In Proceedings of Third International Symposium on Communication Systems Networks and Digital Signal Processing 15(17), 109–112 (2002)
10. Kleinrock, L.: Queuing Systems. Theory, vol. 1. John Wiley & Sons (1975)
11. Sergeev, A., Shanyazov, R., Ilina, D., Bashun, V.: The evaluation of gain of statistical modulation method on the example of qam16 for input data with exponential distribution. In: 2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). pp. 1–5. IEEE (2018)
12. Internet Engineering Task Force (IETF): Guidelines for Considering New Performance Metric Development. `https://tools.ietf.org/html/rfc6390` (2011), [Online; accessed 21-January-2019]
13. Ukhanova, A., Sergeev, A., Forchhammer, S.: Extending jpeg-ls for low-complexity scalable video coding. In: Image Processing: Algorithms and Systems (2011)
14. Sergeev, A.: Signal and code model for hidden data transmission over the voice channel. In: Control and Communications, 2003. SIBCON 2003. The IEEE-Siberian Conference on. pp. 123–126. IEEE (2003)
15. Suznjevic, M., Saldana, J.: Delay limits for real-time services (2016)
16. ITU-T Technology: Watch Report. The Tactile Internet. `https://www.itu.int/dms_pub/itu-t/opb/gen/T-GEN-TWATCH-2014-1-PDF-E.pdf` (2014), [Online; accessed 21-January-2019]
17. Brosh, E., Baset, S., Misra, V., Rubenstein, D., Schulzrinne, H.: The delay-friendliness of tcp for real-time traffic. IEEE/ACM Transactions on Networking (TON) pp. 1478–1491 (2010)
18. ETSI: LTE; Mission Critical Video over LTE (3GPP TS 22.281 version 15.1.0 Release 15). `https://www.etsi.org/deliver/etsi_ts/122200_122299/122281/15.01.00_60/ts_122281v150100p.pdf` (2018), [Online; accessed 21-January-2019]
19. David, H.: Order statistics. John Wiley & Sons (1981)